

October 1, 1998

Alternative GMM Methods for Nonlinear Panel Data Models

Jörg Breitung

Humboldt-Universität Berlin,
Institut für Statistik und Ökonometrie,
Spandauer Strasse 1,
D-10178 Berlin
Email: breitung@wiwi.hu-berlin.de

Michael Lechner

Universität Mannheim,
Institut für Volkswirtschaftslehre und Statistik,
D-68131 Mannheim
Email: Michael.Lechner@vwl.uni-mannheim.de

* Forthcoming in: Matyas, L. (ed.), *Generalised Method of Moments Estimation*, Cambridge University Press, 1998. The research for this paper was carried out within Sonderforschungsbereich 373 at the Humboldt University Berlin and was printed using funds made available by the Deutsche Forschungsgemeinschaft.

1 Introduction

In recent years the GMM approach became increasingly popular for the analysis of panel data (e.g. Avery, Hansen and Hotz [1983], Arrelano and Bond [1991], Keane [1989], Lechner and Breitung [1996]). Combining popular nonlinear models used in microeconomic applications with typical panel data features like an error component structure yields complex models which are complicated or even intractable to be estimated by maximum likelihood. In such cases the GMM approach is an attractive alternative.

A well known example is the probit model, which is one of the working horses whenever models with binary dependent variables are analyzed. Although the nonrobustness of the probit estimates to the model's tight statistical assumption is widely acknowledged, the ease of computation of the maximum likelihood estimator (MLE) - combined with the availability of specification tests - make it an attractive choice for many empirical studies based on cross-sectional data. The panel data version of the probit model allows for a serial correlation of the errors in the latent equations. The problem with such a specification is, however, that the MLE becomes more complicated as in the case of uncorrelated errors.

Two ways to deal with that sort of general problems have emerged in the literature. The idea of simulated maximum likelihood is to find an estimator that approximates the MLE closely, thus retaining the asymptotic efficiency property of the exact MLE. The idea is to use stochastic simulation procedures to obtain approximate choice probabilities (see e.g. Börsch-Supan and Hajivassiliou [1993], or Hajivassiliou, McFadden and Ruud [1996]). The problem with these methods is that they can be very computer intensive, and it may still be difficult to estimate all parameters of the covariance matrix jointly with the regression coefficients.

An alternative approach sacrifices some of the asymptotic efficiency in order to obtain a simple GMM estimator. Since GMM estimators are consistent in the case of serially correlated errors, it is then not necessary to obtain joint estimates of the covariance parameters and the regression coefficients. These estimators are based on the observation that a panel probit model implies a simple probit model when taking each period separately. Therefore, simple moment conditions can be derived from the individual cross-sections and asymptotic theory can be used to minimize the efficiency loss implied by such a procedure. Examples for this kind of estimators can be found in Avery et al. [1983] and Chamberlain [1980, 1984].

In a number of recent papers various other GMM estimators based on these ideas are suggested and compared asymptotically and by means of Monte Carlo simulations (Breitung and Lechner [1996], Bertschek and Lechner (1998), and Lechner and Breitung [1996]). The results of these studies are quite promising: The appropriate variant from the class of GMM estimators provides an estimation procedure that is robust, flexible, easy and fast to compute, and results in a small (in some case negligible) efficiency loss compared to full information maximum likelihood.

In this chapter we review some earlier work on GMM estimation of nonlinear panel data models and suggest some new estimators. In particular we consider the joint estimation of mean and covariance parameters. The previous work focus on first order moment conditions, i.e., the GMM procedure exploits restrictions on the conditional mean while we will consider restrictions on higher moments as well.

The chapter is organized as follows. Section 2 defines the nonlinear panel data model and gives some examples. Section 3 sketch the earlier work concerning the GMM estimation for the conditional mean parameters. Section 4 consider higher order moments for estimating the complete parameter vector and section 5 applies the new approach of Gallant and Tauchen [1996] to select appropriate moment conditions. A minimum distance version of the resulting estimator is suggested in section 6. Section 7 presents the results of a limited Monte Carlo simulation and in section 8 the estimation procedures are applied to an empirical example. Section 9 suggests some conclusions.

2 A Class of Nonlinear Panel Data Models

Let y_{it} be an $m \times 1$ vector of jointly dependent variables and x_{it} is a $k \times 1$ vector of exogenous variables. The indices $i = 1, \dots, N$ and $t = 1, \dots, T$ indicate the cross section unit and the time period of the observation. It is convenient to stack the observations into matrices such that $Y_i = [y_{i1}, \dots, y_{iT}]'$ and $X_i = [x_{i1}, \dots, x_{iT}]'$. As usual in a panel data framework we consider the asymptotic properties for T fixed and $N \rightarrow \infty$.

For the time period t , the nonlinear model is characterized by its conditional density function $h(y_{it}|x_{it}; \theta_0)$ (cf Gouriéroux [1996]). Defining the conditional mean function as $E(y_{it}|x_{it}) = \mu_1(x_{it}; \theta_0)$, the model may be rewritten as

$$y_{it} = \mu_1(x_{it}; \theta_0) + v_{it} , \quad (1)$$

where $E(v_{it}|x_{it}) = 0$ and $E(v_{it}v_{jt}|x_{it}, x_{jt}) = 0$ for $i \neq j$.

EXAMPLE 1 (BINARY CHOICE)

Let $y_{it}^* = x_{it}'\beta_0 + \epsilon_{it}$, where $\epsilon_i = [\epsilon_{i1}, \dots, \epsilon_{iT}]'$ is i.i.d. and ϵ_{it} has a constant marginal distribution function $F_\epsilon(z)$. If $y_{it}^* > 0$, we observe $y_{it} = 1$. Otherwise we have $y_{it} = 0$. The conditional mean function for this model is given by

$$E(y_{it}|x_{it}) = \mu_1(x_{it}; \theta_0) = F_\epsilon(x_{it}'\beta_0).$$

EXAMPLE 2 (MULTINOMIAL LOGIT MODEL)

Let $P_{j,it}$ denote the probability of choosing alternative $j = 0, 1, \dots, J$ given by

$$P_{j,it} = \frac{\exp(x'_{it}\beta_0^j)}{1 + \sum_{j=1}^J \exp(x'_{it}\beta_0^j)},$$

where $\beta_0^0 = 0$. We define J indicator variables¹ $y_{j,it}$ ($j = 1, \dots, J$) which takes a value of one if alternative j is chosen by individual i at time period t and zero otherwise. The conditional mean function for $y_{it} = [y_{1,it}, \dots, y_{J,it}]'$ is

$$E(y_{it}|x_{it}) = \mu_1(x_{it}; \theta_0) = \begin{bmatrix} P_{1,it} \\ \vdots \\ P_{J,it} \end{bmatrix}$$

EXAMPLE 3 (POISSON MODEL)

Let y_{it} be a integer valued random variable drawn from a Poisson distribution with conditional mean function $\mu_1(x_{it}; \theta_0) = \exp(x'_{it}\beta_0)$.

In the examples we give the conditional mean functions for an individual i at time t . Stacking the means of different time periods into a $T \times m$ matrix gives

$$\mu_1(X_i; \theta_0) = \begin{bmatrix} \mu_1(x_{i1}; \theta_0)' \\ \vdots \\ \mu_1(x_{iT}; \theta_0)' \end{bmatrix}.$$

We may include lagged or lead values of the exogenous variables in x_{it} , so that the conditional mean may depend on the exogenous variables from other time periods. Moreover, we may include weakly exogenous variables in the sense that x_{it} is uncorrelated with v_{it} but may be correlated with v_{is} for $s \neq t$. Accordingly, the model may include weakly or strongly exogenous variables. However, it is important to notice that the mean function is the same for all cross section units. This excludes models with heterogeneity in the mean such as fixed effects models.²

Unobserved heterogeneity may be represented by including an individual specific random effect α_i . For example, a random component version of the binary choice model may be constructed as in the following example:

¹Note that the moment condition for $y_{0,it}$ is redundant because the probabilities of the choices sum up to one.

²The reason is that in nonlinear models it is not easy to deal with heterogeneity in the mean. For discrete choice models there are some special cases allowing for fixed effects estimation such as count data models and conditional binary logit models. If the latent dependent variable is partly observable, as in censored or truncated regression models, semiparametric methods may be an attractive alternative (cf Honoré [1992, 1993].)

EXAMPLE 1A (ERROR COMPONENT PROBIT MODEL)

Let y_{it}^* be a latent variable given by

$$y_{it}^* = x_{it}'\beta_0 + \alpha_i + \epsilon_{it},$$

where x_{it} is a vector of strongly exogenous variables, α_i and ϵ_{it} are mutually and serially uncorrelated random variables distributed as $\alpha_i \sim \mathcal{N}(0, \sigma_\alpha^2)$ and $\epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, where we normalize the variances as $\sigma_\epsilon^2 + \sigma_\alpha^2 = 1$. The observed dependent variable y_{it} is one if $y_{it}^* > 0$ and zero otherwise. The parameter vector of the model is $\theta = [\beta_0', \rho_0]'$, where $\rho_0 = \sigma_\alpha^2/(\sigma_\epsilon^2 + \sigma_\alpha^2)$ denotes the serial correlation and the mean function is identical to the one of Example 1, where $F_\epsilon(z)$ is the standard normal c.d.f. This error component probit model is a special case of Example 1 and will serve as the leading example in what follows.

3 GMM estimators for the conditional mean

Assume that we are interested in the conditional mean function given by $\mu_1(X_i; \theta)$. In many applications, the mean function does not depend on the complete parameter vector. In this case we may write $\mu_1(X_i; \theta) = \mu_1(X_i; \theta^{(1)})$, where $\theta^{(1)}$ is a $p_1 \times 1$ subvector of θ . The remaining parameters are treated as nuisance parameters. For example, in the error components probit model given in Example 1a we may be interested in β_0 but not in the correlation coefficient ρ_0 . Accordingly, it is convenient to focus on the conditional mean when constructing a GMM estimator. We define first order moments as

$$f_1(y_i, X_i; \theta^{(1)}) = y_i - \mu_1(X_i; \theta^{(1)}) \quad (2)$$

with the moment condition

$$E[f_1(X_i; \theta_0^{(1)})|X_i] = 0. \quad (3)$$

Using the law of iterated expectations the conditional moment restrictions can be expressed as a set of unconditional moment restrictions

$$E[B(X_i)f_1(y_i, X_i; \theta_0^{(1)})] = 0, \quad (4)$$

where $B(X_i)$ is a $p_1 \times T$ matrix of functions on X_i . The optimal choice of $B(X_i)$ is (cf Newey [1993])

$$B^*(X_i) = C \cdot D(X_i; \theta_0^{(1)})'\Omega(X_i; \theta_0)^{-1}, \quad (5)$$

where

$$\begin{aligned} D(X_i; \theta^{(1)}) &= E[\partial f_1(y_i, X_i; \theta^{(1)})/\partial \theta^{(1)}|X_i] \\ \Omega(X_i; \theta) &= E[(f_1(y_i, X_i; \theta^{(1)}))f_1(y_i, X_i; \theta^{(1)})'|X_i] \end{aligned}$$

and C is some nonsingular squared matrix.

For the panel probit model (Example 1a) we obtain:

$$D(X_i; \beta) = \begin{bmatrix} -\phi(x'_{i1}\beta_0)x'_{i1} \\ \vdots \\ -\phi(x'_{iT}\beta_0)x'_{iT} \end{bmatrix}$$

$$\Omega(X_i; \beta, \rho) = (\omega_{ts,i})$$

$$\omega_{ts,i}(x_{it}, x_{is}; \beta, \rho) = \begin{cases} \Phi(x'_{it}\beta)[1 - \Phi(x'_{it}\beta)] & \text{for } t = s \\ \Phi^{(2)}(x'_{it}\beta, x'_{is}\beta, \rho) - \Phi(x'_{it}\beta)\Phi(x'_{is}\beta) & \text{otherwise,} \end{cases}$$

where $\phi(x'_{it}\beta)$ denotes the p.d.f. of the univariate standard normal distribution and $\Phi^{(2)}(x'_{it}\beta, x'_{is}\beta, \rho)$ indicates a bivariate normal c.d.f. with correlation ρ .

It is important to note that $D(X_i; \beta)$ does not depend on the covariance parameter ρ , whereas the computation of $\omega_{ts,i}$ requires the value of ρ . Accordingly, for the asymptotically optimal instruments we need estimates of the mean and the covariance parameters. There are several strategies to deal with this problem. First, ρ may be fixed at some arbitrary level, say $\rho = 0$. This gives a computationally simple estimator that is consistent but inefficient. The second possibility is to apply some rough approximation such as the “small sigma approximation” (cf. Breitung and Lechner [1996]). Again, an efficiency loss may result from this approximation.

The third possibility is to estimate $T(T-1)/2$ bivariate probit models to obtain direct estimates. This approach is asymptotically efficient, but fairly time consuming. In particular, convergence problems may occur, as it often happens in fairly small sample sizes. Finally, Bertschek and Lechner [1998] suggest the use of nonparametric methods, (e.g. the k-nearest neighbor method) to avoid the estimation of the correlation coefficient. Based on an empirical application and an extensive Monte Carlo study they find that the nonparametric GMM estimator approach the efficiency of the MLE.

Another possibility is to use a high dimensional simple function such as

$$B(X_i) = I_T \otimes \text{vec}(X_i)'$$

or

$$B(X_i) = \begin{bmatrix} x_{i1}/s_{i1} & 0 & \dots & 0 \\ 0 & x_{i2}/s_{i2} & & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \dots & x_{iT}/s_{iT} \end{bmatrix},$$

where $s_{it} = \Phi(x'_{it}\beta)[1 - \Phi(x'_{it}\beta)]$ is the inverse of the error variance (see Avery et al. [1983], Breitung and Lechner [1996]). The main problem with this approach is that a large number of moment conditions is needed to approach the efficient GMM estimator. In finite samples, however, GMM estimators perform poorly

if the number of moment conditions gets large relative to the sample size (e.g. Breitung and Lechner [1996]).

4 Higher Order Moment Conditions

So far we confined ourselves to the moment restrictions for the conditional mean. Accordingly, this approach does not render estimates for other parameters like the conditional variance. Moreover, the efficiency of the GMM procedure may be improved by considering higher order moment conditions. If y_i is univariate (as for the cross section probit model, for example) we may define moments of degree k as

$$f_k(y_i, x_i; \theta) = [y_i - \mu_1(x_i; \theta)]^k - \mu_k(x_i; \theta) \quad (6)$$

where $\mu_k = E\{[y_i - \mu_1(X_i; \theta)]^k | x_i\}$. The corresponding moment condition is

$$E[f_k(y_i, x_i; \theta_0) | x_i] = 0 . \quad (7)$$

If y_i is a vector, all relevant moments are stacked into an appropriate vector. For the panel probit model the typical element for a vector with $k = 2$ is given by

$$f_2^{(ts)}(y_i, X_i; \beta, \rho) = [y_{it} - \Phi(x'_{it}\beta)][y_{is} - \Phi(x'_{is}\beta)] - \omega_{ts,i}(x_{it}, x_{is}; \beta, \rho), \quad (8)$$

where $\omega_{ts,i}$ is defined as in (6). For $t \neq s$, the moment condition requires the evaluation of the bivariate normal c.d.f. and is, therefore, more complicated than imposing just the conditional mean restrictions.

Intuitively, as we include an increasing number of moment conditions, the moment conditions will give an accurate characterization of the conditional distribution and, therefore, the GMM estimator will tend to the MLE. There are, however, serious problems with such an approach. First, higher order moment conditions easily become quite complicated so that the GMM procedure may even be more burdensome than the corresponding MLE. Second, the number of moment conditions increase rapidly with k . It is therefore desirable to have an alternative method for generating moment conditions rather than considering higher order conditional moments.

5 Selecting Moment Conditions: The Gallant-Tauchen Approach

Gallant and Tauchen [1996] suggest to employ an auxiliary (possibly misspecified) model for generating moment conditions from the scores of the pseudo MLE. The idea behind this approach is that the scores of an accurate representation of the main features of the model (the score generator) may provide efficient moment conditions. In fact, if the auxiliary model “smoothly embed”

the structural model, then the GMM estimator derived from the score generator is asymptotically efficient (Gallant and Tauchen [1996]).

Consider the panel probit model given in Example 1a. Using a linear error component model $y_{it} = x'_{it}\gamma + \alpha_i^* + \epsilon_{it}^*$ as auxiliary model gives rise to the following moments³

$$s_{N,1}(\lambda) = \frac{1}{\sigma_{\epsilon^*}^2 NT} \sum_{t=1}^T \sum_{i=1}^N (u_{it} - \psi \bar{u}_i) x_{it} \quad (9)$$

$$s_{N,2}(\lambda) = \frac{1}{N} \sum_{i=1}^N \bar{u}_i^2 - \sigma_{\alpha^*}^2 - \frac{1}{T} \sigma_{\epsilon^*}^2 \quad (10)$$

$$s_{3,N}(\lambda) = \frac{1}{N(T-1)} \sum_{t=1}^T \sum_{i=1}^N (u_{it} - \bar{u}_i)^2 - \sigma_{\alpha^*}^2, \quad (11)$$

where $\lambda = [\gamma', \sigma_{\alpha^*}^2, \sigma_{\epsilon^*}^2]'$, $u_{it} = y_{it} - x'_{it}\gamma$, $\bar{u}_i = T^{-1} \sum_t u_{it}$, and $\psi = T\sigma_{\alpha^*}^2 / (T\sigma_{\alpha^*}^2 + \sigma_{\epsilon^*}^2)$. It is important to note that the variance of the errors in a probit model is not identified and is therefore set to one. As a consequence the moment $s_{3,N}(\lambda)$ is dropped to obtain a nonsingular covariance matrix of the conditional moments.

Although the linear model is misspecified, it may approximate the crucial features of the underlying nonlinear model. Whenever it is possible to derive the relationship between the parameters of the structural and auxiliary model, we are able to compute estimates for the structural model from the estimated auxiliary model.

Let θ denote the vector of structural parameters, (i.e. $\theta = [\beta', \rho]'$ in our example) and

$$f_N(\theta, \lambda) = E_\theta[s_N(\lambda)], \quad (12)$$

denotes the expected scores, where $s_N(\lambda) = [s_{1,N}(\lambda), s_{2,N}(\lambda), s_{3,N}(\lambda)]'$ and E_θ indicates the expectation with respect to the structural model given the parameter vector θ . Then, Gallant and Tauchen [1996] suggest to estimate θ by minimizing the objective function

$$\hat{\theta}_{GT} = \underset{\theta}{\operatorname{argmin}} \{f_N(\theta, \tilde{\lambda})' A_N f_N(\theta, \tilde{\lambda})\}, \quad (13)$$

where $\tilde{\lambda}$ is the pseudo MLE given by $s_N(\tilde{\lambda}) = 0$ and A_N tends to the optimal weight matrix as $N \rightarrow \infty$. If the dimension of $f_N(\theta, \tilde{\lambda})$ is the same as the number of structural parameters, we can compute $\hat{\theta}_{GT}$ from solving $f_N(\hat{\theta}_{GT}, \tilde{\lambda}) = 0$.

In our example, the moments $f_N(\theta, \tilde{\lambda})$ can be computed by using the bivariate normal c.d.f. (see section 3). Another possibility, which is in particular useful for more complicated models, is to approximate (12) using Monte Carlo techniques (see, e.g., Gouriéroux and Monfort [1993] and Gallant and Tauchen [1996]).

³These expressions are derived from the scores given in Hsiao (1986, p.39).

To compute the GMM estimator we need (estimates for) the derivatives

$$F_N(\theta) = \frac{\partial E_\theta \{f_N[\theta, \tilde{\lambda}]\}}{\partial \theta'}, \quad (14)$$

which is a quite complicated function in general. It is important to notice that in (12) the expectation is computed by treating λ as given while in (14) $\tilde{\lambda}$ is a random variable depending on θ . In practice, this expression is estimated using simulation methods.

Usually, the variance of the GMM estimator with optimal weight matrix is estimated as

$$\hat{V}_N = [F_N(\hat{\theta})' A_N F_N(\hat{\theta})]^{-1}.$$

However, for the consistency of this estimator it is required that the auxiliary model is correctly specified so that there exists a value λ^* such that the “score contributions” satisfy

$$E_\theta[s(y_i, X_i; \lambda^*)] = 0 \quad \text{for all } i = 1, \dots, N,$$

where $s_N(\lambda) = \sum_i s(y_i, X_i; \lambda)$. For misspecified auxiliary models we therefore need to apply a different estimator for the covariance matrix of $\hat{\theta}$ given by

$$\bar{V}_N = [F_N(\hat{\theta})' A_N F_N(\hat{\theta})]^{-1} F_N(\hat{\theta})' A_N \bar{\Sigma}_f A_N F_N(\hat{\theta}) [F_N(\hat{\theta})' A_N F_N(\hat{\theta})]^{-1}$$

where

$$\bar{\Sigma}_f = \frac{1}{N} \sum_{i=1}^N [s(y_i, X_i; \tilde{\lambda}) - m_i(\tilde{\lambda})][s(y_i, X_i; \tilde{\lambda}) - m_i(\tilde{\lambda})]' \quad (15)$$

and

$$m_i(\lambda) = E_\theta[s(y_i, X_i; \lambda)].$$

Obviously, this modification, suggested by Gallant and Tauchen [1996] implies a considerable increase in computational burden compared with the conventional estimator.

In sum, although the approach suggested by Gallant and Tauchen [1996] is attractive for selecting suitable moment conditions, it implies a great deal of computational effort. Hence, this approach is not recommended for rather simple models considered here for the ease of exposition. If the model becomes much more complex and no convenient expressions for the first moments are available, the Gallant-Tauchen approach seem to be an attractive devise to select useful moments for the GMM procedure.

6 A Minimum Distance Approach

It is possible to construct a computationally more convenient minimum distance procedure with the same asymptotic properties as the Gallant-Tauchen estimator. This approach was suggested by Gouriéroux, Monfort and Renault [1993].

Applying a Taylor series expansion to the moments derived from the auxiliary model gives

$$\begin{aligned} f_N(\theta, \tilde{\lambda}) &= E_\theta[s_N(\tilde{\lambda})] \\ &= E_\theta[s_N(\lambda^*)] + H_N(\lambda^*)(\tilde{\lambda} - \lambda^*) + o_p(T^{-1/2}) \\ &= H_N(\lambda^*)(\tilde{\lambda} - \lambda^*) + o_p(T^{-1/2}) \end{aligned}$$

where $H_N(\lambda) = \partial E_\theta[s_N(\lambda)] / \partial \lambda'$ and λ^* is the value of λ such that $E_\theta[s_N(\lambda^*)] = 0$. As the sample size tends to infinity, θ^* converges to the “pseudo-true value” defined in White [1982]. If $H_N(\lambda^*)$ is a regular matrix, then minimizing $f_N(\theta, \tilde{\lambda})' A_N f_N(\theta, \tilde{\lambda})$ is asymptotically equivalent to

$$\hat{\theta}_{MD} = \operatorname{argmin}\{(\tilde{\lambda} - \lambda^*)' V_{\tilde{\lambda}}(\tilde{\lambda} - \lambda^*)\},$$

where $V_{\tilde{\lambda}} = H_N(\lambda^*)' A_N H_N(\lambda^*)$ tends to the covariance matrix of $\tilde{\lambda} - \lambda^*$. For convenience the notation does not make explicit the dependence of λ^* on θ .

For computing $\hat{\theta}_{MD}$ the derivative

$$D_\lambda^*(\theta) = \frac{\partial \lambda^*}{\partial \theta'}$$

needs to be evaluated. Usually, this derivative is much easier to compute than the derivative $F_N(\theta)$ needed to compute the Gallant-Tauchen estimator. Since $E_\theta(\tilde{\theta})$ tends to θ^* as $N \rightarrow \infty$, this derivative is asymptotically equivalent to

$$\widetilde{D}_\lambda(\theta) = \frac{\partial E_\theta(\tilde{\lambda})}{\partial \theta'},$$

which can easily be estimated by simulation techniques.

A second important advantage of this variant of the Gallant-Tauchen approach is that the covariance matrix $\Sigma_{MD} = E[(\hat{\theta}_{MD} - \theta)(\hat{\theta}_{MD} - \theta)']$ can be estimated as

$$\tilde{\Sigma}_{MD} = \widetilde{D}_\lambda(\theta)' \tilde{\Sigma}_\lambda \widetilde{D}_\lambda(\theta)$$

where $\tilde{\Sigma}_\lambda = E_\theta[(\tilde{\lambda} - \lambda^*)(\tilde{\lambda} - \lambda^*)']$, which is much easier to estimate than the expression (15).

7 Finite Sample Properties

7.1 Data Generating Process

To compare the small sample properties of different GMM estimators, we simulate data according to the following model:

$$\begin{aligned}
 y_{it} &= \mathbb{I}(\beta^C + \beta^D x_{it}^D + \beta^N x_{it}^N + u_{it} > 0) \\
 x_{it}^D &= \mathbb{I}(\tilde{x}_{it}^D > 0), & P(\tilde{x}_{it}^D > 0) &= 0.5 \\
 x_{it}^N &= 0.5x_{i,t-1}^N + 0.05t + \eta_{it}, & \eta_{it} &\sim U[-1, 1], \\
 u_{it} &= \delta c_i + \epsilon_{it}, & c_i &\sim N(0, 1), \\
 \epsilon_{it} &= \alpha \epsilon_{i,t-1} + \sigma \tilde{\epsilon}_{it}, & \tilde{\epsilon}_{it} &\sim N(0, 1), \\
 i &= 1, \dots, N, \quad t = 1, \dots, T.
 \end{aligned}$$

which is also used in Breitung and Lechner [1996] and Bertschek and Lechner [1998]. The parameters $(\beta^C, \beta^D, \beta^N, \delta, \alpha, \sigma)$ are fixed coefficients and $\mathbb{I}(\cdot)$ is an indicator function, which is one if its argument is true and zero otherwise. The parameter σ is chosen such that the variance of u_{it} is unity. All random numbers are drawn independently over time and individuals. The first regressor is a serially uncorrelated indicator variable, whereas the second regressor is a smooth variable with bounded support. The dependence on lagged values and on a time trend induces a correlation over time. This type of regressor has been suggested by Nerlove [1971] and was also used for example by Heckman [1981].

We set $\beta^C = -0.75$, $\beta^D = \beta^N = 1$ in all simulations and $T^{-1} \sum_{t=1}^T \text{var}(u_{it}) = 1$. To represent typical sample sizes encountered in empirical applications we let $N = 100, 400$ and $T = 5, 10$. Depending on the DGP, 500 or 1000 replications (R) were generated. In order to diminish the impact of initial conditions, the dynamic processes have been started at $t = -10$ with $x_{i,t-11}^N = \epsilon_{i,t-11} = 0$.

In the simulations two different specifications are considered. First, a pure error component model with serially and mutually uncorrelated error components is obtained by setting $\alpha = 0$ and $\delta = \sqrt{0.5}$. Furthermore $\sigma = \sqrt{0.5}$ so that $E(u_{it}) = 1$. The correlation coefficient results as $\rho = 0.5$.

The second specification removes the equi-correlation pattern by setting $\alpha = 0.8$, $\delta = 0.2$ and $\sigma = 0.5$. In such a specification, the serial correlation is persistent but declines with an increasing lag length. The maximum correlation coefficient is 0.8 for a single lag and the correlation decrease to 0.4 for a lag length of four (the maximum lag length when letting $T = 5$).

7.2 Estimators

The first estimator is the MLE computed as in Butler and Moffitt [1982]. The number of evaluation points is set to 5 as a compromise between computational speed and numerical accuracy. The robust estimator according to White [1982] is used to estimate the standard errors. The results for this estimator are indicated

by the acronym ML-RE. The pooled estimator denotes the ML estimator ignoring the panel structure of the data. The standard errors are estimated allowing for serial correlation as in Avery et al. [1983].

The infeasible GMM-IV estimator is the optimal GMM estimator using the conditional mean restrictions (see Section 3). To compute the optimal instrument matrix the true correlation ρ_0 is used. Bertschek and Lechner [1998] propose a feasible GMM estimator that estimates the unknown quantities in the expression for the optimal instruments by nonparametric methods. The version that performs best in their Monte Carlo study is labeled GMM-WNP.

Another approach to obtain the optimal instruments is to get a consistent estimate of the unknown correlation coefficient. The estimator GMM-IV(param) proposed here consists of three steps. In the first step a pooled probit model is estimated to obtain consistent estimates β . Then $T(T-1)/2$ second moments as in eq. (8) are used to compute consistent estimates of the correlations. For given values of β , each moment condition depends only on the unknown correlation coefficient, which is bounded between -1 and $+1$. Hence, grid search methods are used to determine the correlation coefficients. For a random effects model, using one such moment condition is sufficient for obtaining a consistent estimate of ρ . However, our estimation procedure allows for different values of the $T(T-1)/2$ correlation coefficients ρ_{ts} ($t, s = 1, \dots, T$). This estimator is still optimal even when the covariance structure is more general than the equi-correlation structure of a random effects model.

The pooled probit estimator as well as the GMM estimators with asymptotically optimal instruments are consistent no matter of the true error correlation. Furthermore, all GMM estimators have the same limiting distribution whether the true or consistently estimated optimal instruments are used. To yield a nonparametric estimate of the optimal instruments, a nearest-neighbor approach is applied (see Bertschek and Lechner [1998] for details). This resulting estimator is labeled as GMM-WNP.

Following Gallant and Tauchen [1996] we employ simple auxiliary models as score generators. First we use a linear error components model with scores given in (9) and (10). To compute the expectation of the scores $f_N(\theta, \tilde{\lambda})$ we generate 100 replications of the model and compute the average of the scores. Using a Taylor expansion

$$f_N(\theta + \delta, \tilde{\lambda}) = f_N(\theta, \tilde{\lambda}) + \frac{\partial f_N(\theta, \tilde{\lambda})}{\partial \theta'} \delta + \tilde{r}, \quad (16)$$

where \tilde{r} is a remainder term of order $O(\delta^2)$, the derivatives are estimated by a least squares regression of $f_N(\theta + \delta, \tilde{\lambda}) - f_N(\theta, \tilde{\lambda})$ on δ , where δ is a vector of normally distributed random numbers with zero mean and $E(\delta^2) = 0.02$. We use 100 realizations of δ for computing the regression lines. This regression estimator has the advantage that it does not suffer from problems due to the discontinuity and non-differentiability of the simulator. At every iteration step

of θ , the same sequence of random numbers are used. Using the *pooled* estimator with an estimate of ρ based on the small-sigma estimate (Breitung and Lechner [1996]) as initial values, the algorithm usually converges after 5-10 iterations. The resulting estimates are labeled as *GT-score*.

Our practical experiences with such an algorithm suggest that the convergence properties crucially depend on the number of replications used to estimate the conditional mean and derivatives of the moments. In particular, the least-squares estimate of the derivatives requires at least 100 replications to obtain reliable estimates of the gradients. Furthermore, the step-length δ in the Taylor expansion should be small enough to avoid a substantial bias but must be large enough to achieve acceptable properties of the least-square estimator. In our simulations we found that a value of $E(\delta^2) = 0.02$ provides a reasonable trade-off for our data generating process. For other processes, however, suitable values for δ or the number of replications may be different.

Adding the scores from the (cross section) probit estimator applied to the pooled dataset we obtain three additional moments. The weight matrix are computed using the *GT-score* estimator. The computational details for the simulations are the same as for the *GT-score* procedure. The resulting estimator is labeled as *GT-score*⁺. Since the estimator employ the instruments of the pooled probit estimator, it is asymptotically more efficient as the *pooled* estimator.

For both estimators using the Gallant-Tauchen approach the respective minimum distance estimators are computed. Accordingly, we denote these estimators as *GT-MD* and *GT-MD*⁺. We use 300 Monte Carlo realizations of the model for computing the moments. The matrix of derivatives $\widetilde{D}_\lambda(\theta)$ were computed by least-squares using a Taylor expansion of $\widetilde{\lambda}$.

To compare the performance of the estimators we compute the root mean square error (RMSE) and the median absolute error (MAE). For the estimated standard errors of the coefficient we compute the relative bias. The precise definitions of these measures are given in Table 1.

7.3 Results

Table 2 presents the results for the specification with pure random effects. It turns out that with respect to RMSE and MAE the MLE performs best among all estimators. The second best estimator is the GMM estimator based on the optimal instruments derived from the conditional mean restrictions (*Infeasible GMM*). However, this estimator is based on a known correlation parameter ρ_0 so that such an estimator is of limited use in practice.

With respect to the other GMM estimators, the ranking is not as clear. Generally, the GMM procedures using information about the error correlation like the feasible versions of *GMM-IV* or the estimators using a linear error component model as a score generator perform better than the *pooled* estimator ignoring the error correlation altogether.

Comparing the small sample properties of the two asymptotically equivalent GMM estimators $GMM-IV(param)$ and $GMM-WNP$, the latter estimator appears to be superior for all DGP's, with the exception of the random effects DGP with $N = 400$ and $T = 5$. The potential small sample problems of these estimators is related to the estimation of the inverse of the conditional covariance matrix of the residuals for each individual. $GMM-WNP$ uses nonparametric methods that performs very well even in fairly samples (see Bertschek and Lechner [1998] for details). The problem with $GMM-IV(param)$ is that some of the estimated ρ_{ts} coefficient may end up at the boundary of the parameter space $[-1, +1]$ when N is 'not large enough'. This is a particular problem when the number of coefficients to be estimated gets large (45 in the case of $T = 10$). A potential remedy is to enforce equality of the ρ_{ts} in the second step of the estimation. There are however two drawbacks of such a procedure: First, the estimator is no longer asymptotically efficient when the true DGP is different from the random effect model. Second, the simplicity of the estimator is lost. Therefore, we conclude that in applications $GMM-IV(param)$ may be a preferable option only if an estimate of the correlation structure of latent residuals is of interest and if the dimension N is sufficiently large relative to the dimension T .

The Gallant-Tauchen estimators derived from a linear error component model seem to work well. This is perhaps surprising since the linear model is quite a crude approximation to the panel probit model. In fact, there is much room for improving the fit of the auxiliary model. For example, the nonlinear mean function and heteroscedasticity of the errors are important features which are neglected by the linear approximation. Nevertheless, the scores of the linear model obviously provide useful moment conditions to be exploited by a GMM procedure.

In small samples, the original version and the minimum distance variant perform somewhat differently. For the smaller set of moment conditions the Gallant-Tauchen GMM estimator ($GT-Score$) outperforms the respective minimum distance estimator ($GT-MD$), while for the enhanced set of moment conditions the ($GT-MD+$) estimator performs better than the ($GT-Score+$) estimator. In all, however, the differences are small relative to the simulation error.

The estimation of the standard errors for the coefficients are extremely biased for Gallant-Tauchen estimators. The reason is that the standard errors are estimated assuming a correctly specified auxiliary model. Of course, this is not true in our application and it turns out that the resulting bias can be immense. Unfortunately, the computational effort for correcting the estimates along the lines suggested by Gallant and Tauchen [1996] was beyond the time schedule for the present work.

The problems with the estimation of the standard errors can be side-stepped by using the minimum distance approach. In fact, the estimation of the standard errors for the ($GT-MD+$) procedure seems to perform acceptable. However, the standard errors for the ($GT-MD$) estimator still possess a substantial bias.

To study the performance of the estimators under more general conditions, we introduce a persistent autocorrelation in addition to the random effects. Obviously, such a data generating process is difficult to distinguish empirically in a sample with a small number of time periods. All GMM estimators designed for the error component model remain consistent in the presence of a more general form of serial correlation. Thus, it is interesting to know whether the efficiency ranking is robust to differences in the autocorrelation pattern. Table 3 presents the results for such a process.

The conclusions from simulations with other sample sizes are qualitative similar. It turns out that the relative performance of the estimators is roughly similar to the case of a pure random effects model. However, it appears that the *ML-RE* estimator loses most of its relative advantages. Furthermore, the Gallant-Tauchen type of estimators perform worse than the competitors based on the conditional mean function. On the other hand, the *GMM-WNP* estimator turns out to have the most attractive properties. It is simple to compute and has favorable small sample properties for the DGP considered here.

8 An Application

An empirical example for our discussion of panel probit models is the analysis of firms' innovative activity as a response to imports and foreign direct investment (FDI) as considered in Bertschek [1995]. The main hypothesis put forward in that paper is that imports and inward FDI have positive effects on the innovative activity of domestic firms. The intuition for this effect is that imports and FDI represent a competitive threat to domestic firms. Competition on the domestic market is enhanced and the profitability of the domestic firms might be reduced.

As a consequence, these firms have to produce more efficiently. Increasing the innovative activity is one possibility to react to this competitive threat and to maintain the market position. The dependent variable available in the data takes the value one if a product innovation has been realized within the last year and the value zero otherwise. The binary character of this variable leads us to formulate the model in terms of a latent variable that represents for instance the firms' unobservable expenditures for innovation that is linearly related to the explanatory variables.

The firm-level data have been collected by the Ifo-Institute, Munich ('Ifo-Konjunkturtest') and have been merged with official statistics from the German Statistical Yearbooks. The binary dependent variable indicates whether a firm reports having realized a product innovation within the last year or not. The independent variables refer to the market structure, in particular the market size of the industry $\ln(\text{sales})$, the shares of imports and FDI in the supply on the domestic market *import share* and *FDI-share*, the *productivity* as a measure of the competitiveness of the industry as well as two variables indicating whether a

firm belongs to the *raw materials* or to the *investment goods* industry. Moreover, including the *relative firm size* allows to take account of the innovation — firm size relation often discussed in the literature. Hence, all variables with exception of the firm size are measured at the industry-level (for descriptive statistics see Table 4).

The estimators applied to the example include the simplest one (pooled with GMM standard errors), both feasible GMM estimators based on second order moments and the minimum distance versions of the Gallant-Tauchen estimator. For the latter estimator, we use 1000 Monte Carlo replications for the simulated moments as well as its derivatives. Results for other estimators for that example can be found in Bertschek and Lechner [1998].

The results of the different GMM procedures are presented in Table 5. In all they are quite similar and yield the same conclusions. Both *import share* and *FDI-share* have positive and significant effects on product innovative activity. As expected by the Schumpeterian hypothesis that large firms are more innovative than small firms the *firm size* variable has a positive and significant impact. The coefficient of *productivity* is significantly negative for *pooled*, *GMM-IV(param)* and *GMM-WNP* but insignificant when using the *GT-MD*.

An interesting finding is that the estimated standard errors of *GT-MD* tend to be substantially greater than the corresponding estimates of the alternative estimators. We have tried different numbers of replications or values of δ . The problem is that if δ is a small number, then the simulation error is large in relative terms, while for large values of δ , the estimates of the derivative are biased. The standard errors presented in Table 5 are based on 10.000 replications and $\delta = 0.05 \cdot \hat{\beta}_P$, where $\hat{\beta}_P$ denotes the *pooled* estimator. We decided to choose a relative step size, because the parameter values appears to be quite different in magnitude. Repeating the computation of the standard errors using a different sequence of random numbers shows that these estimates reveal a considerable variability. Hence, these estimates do not seem very reliable and must be interpreted with caution.

It is interesting to consider the serial correlation of the errors between different time periods (see Table 6) They are obtained as by-product in the computation of the second step of *GMM-IV(param)* (see section 7.2). It turns out that the autocorrelation function decays with increasing lag length. This result suggests that an autoregressive pattern is more suitable than the equi-correlation implied by an error component model. In any case, the GMM estimators remain consistent regardless of the actual form of the autocorrelation function.

9 Concluding Remarks

GMM is an attractive approach for estimating complex models like nonlinear error component models popular in current econometric research. Simple estimators

can be constructed by using restrictions implied by the conditional mean function. Asymptotically optimal GMM estimators based on the conditional mean function are obtained by using a parametric or nonparametric approach. Our Monte Carlo results clearly indicate that the nonparametric approach is superior in small samples.

In addition we consider further moment conditions derived from higher order moments or the “score generator” as proposed by Gallant and Tauchen [1996]. From a practical perspective the latter approach is appealing, because it provides simple moment conditions and may yield highly efficient estimators. However, the computational effort is considerable for such estimators. Following Gouriéroux et al. [1993], we adopt a minimum distance analog of the Gallant-Tauchen estimator, which is much simpler to compute and generally renders valid estimates of the standard errors.

Our Monte Carlo results demonstrate that the GMM procedures considered in this paper perform well relative to the MLE. Although the Gallant-Tauchen estimator is based on a very simple “score generator”, the efficiency comes close to the MLE. However, the computational burden of the Gallant-Tauchen approach is immense and there are serious problems when estimating the standard errors for the parameters. Therefore, we do not recommend this estimator for models like the error component probit models, where much simpler GMM procedures with better small sample properties are available. However, if the model is more complicated and simple GMM estimators do not exist, the Gallant-Tauchen approach may be a useful device for providing efficient moment conditions.

References

- Arellano, M. and Bond, S.** [1991]: Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277–297.
- Avery, R., Hansen, L. and Hotz, V.** [1983]: Multiperiod Probit Models and Orthogonality Condition Estimation, *International Economic Review*, 24, 21–35.
- Bertschek, I.** [1995]: Product and Process Innovation as a Response to Increasing Imports and Foreign Direct Investment, *Journal of Industrial Economics*, 43[4]; 341–357.
- Bertschek, I. and Lechner, M.** [1998]: Convenient Estimators for the Panel Probit Model, forthcoming in *Journal of Econometrics*.
- Boersch-Supan, A. and Hajivassiliou, V.A.** [1993]: Smooth Unbiased Multivariate Probabilities Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models, *Journal of Econometrics*, 58, 347–368.

- Breitung, J. and Lechner, M.** [1996]: Estimation de modèles non linéaires sur données de panel par la méthode des moments généralisés, *Economie et Prevision*, 126, 191–204. (The English version is available as SFB discussion paper No. 67 / 1995 , Humboldt-University, Berlin).
- Butler, I.S. and Moffitt, R.** [1982]: A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model, *Econometrica*, 50, 761-764.
- Chamberlain, G.** [1980]: Analysis of Covariance with Qualitative Data, *Review of Economic Studies*, 47, 225-238.
- Chamberlain, G.** [1984]: Panel Data, in Griliches, Z. and Intriligator, M.D. [eds.], *Handbook of Econometrics*, Vol. II, Ch. 22, Amsterdam: North-Holland.
- Gallant, A.R. and Tauchen, G.** [1996]: Which Moments to Match, *Econometric Theory*, 12, 657–681.
- Gouriéroux, C. Monfort, A. and Renault, E.** [1993], Indirect Inference, *Journal of Applied Econometrics*, 8, S85–S118.
- Gouriéroux, C.** [1996]: Introduction to Nonlinear Models, in: L. Mátyás and P. Sevestre [eds], *The Econometrics of Panel Data*, 2. vol., Chap. 15, 399–409, Dordrecht: Kluwer.
- Gourieroux, C. and Monfort, A.** [1993]: Simulation-based Inference: a survey with special reference to panel data models, *Journal of Econometrics*, 59, 5–33.
- Guilkey, D.K. and Murphy, J.L.** [1993]: Estimation and Testing in the Random Effects Probit Model, *Journal of Econometrics*, 59, 301-317.
- Hajivassiliou, V.A.** [1993]: Simulation Estimation Methods for Limited Dependent Variable Models, in: Maddala, G.S., Rao, C.R. and Vinod, H.D. [eds.], *Handbook of Statistics*, Vol. 11: Econometrics, Ch. 19, Amsterdam: North-Holland.
- Hansen, L.P.** [1982]: Large Sample Properties of Generalized Methods of Moments Estimators, *Econometrica*, 50, 1029–1055.
- Heckman, J.J.** [1981]: The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time - Discrete Data Stochastic Process and Some Monte Carlo Evidence, in: Manski, C. and McFadden, D. [eds]: *Structural Analysis of Discrete Data*, Cambridge: MIT-Press.

- Honoré, B.** [1992]: Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects, *Econometrica*, 60, 533–565.
- Honoré, B.** [1993]: Orthogonality Conditions for Tobit Models with Fixed Effects and Lagged Dependent Variables, *Journal of Econometrics*, 59, 35–61.
- Hsiao, C.** [1992]: Logit and Probit Models, in: Matyas, L. and Sevestre, P. [eds.], *The Econometrics of Panel Data*, Ch. 11, Dordrecht: Kluwer.
- Keane, M.P.** [1993]: Simulation Estimation for Panel Data Models with Limited Dependent Variables, in Maddala, G.S., Rao, C.R. and Vinod, H.D. (eds.), *Handbook of Statistics*, Vol. 11: Econometrics, Ch. 20, Amsterdam: North-Holland.
- Keane, M.P.** [1994]: A Computationally Practical Simulation Estimator for Panel Data, *Econometrica*, 62, 95–116.
- Lechner, M. and Breitung, J.** [1996]: Some GMM Estimation Methods and Specification Tests for Nonlinear Models, in: L. Mátyás and P. Sevestre [eds.], *The Econometrics of Panel Data*, 2. vol., Chap. 15, 399–409, Dordrecht: Kluwer.
- McFadden** [1989]: A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, *Econometrica*, 57, 995–1026.
- McFadden, D. and Ruud, P.A.** [1995]: Estimation by Simulation, *The Review of Economics and Statistics*, 76, 591–608.
- Nerlove, M.** (1971): Further Evidence on the Estimation of Dynamic Economic Relations From a Time Series of Cross Sections, *Econometrica*, 39, 359–383.
- Newey, W.K.** [1993]: Efficient Estimation of Models with Conditional Moment Restrictions, in Maddala, G.S., Rao, C.R. and Vinod, H.D. [eds.], *Handbook of Statistics*, Vol. 11: Econometrics, Ch. 16, Amsterdam: North-Holland.
- Newey, W.K. and McFadden, D.L.** [1994]: Large Sample Estimation and Hypothesis Testing, in: Engle, R.F and McFadden, D.L. [eds.], *Handbook of Econometrics*, Vol. IV, Elsevier: North Holland.
- White, H.** [1982]: Maximum Likelihood Estimation of Misspecified Models, *Econometrica*, 50, 1–26.

Table 1: Measures of Accuracy Used in the Monte Carlo Study

RMSE:	root mean square error	$\sqrt{\frac{100}{R} \sum_{r=1}^R (\hat{\theta}_r - \theta_0)^2}$
MAE:	median absolute error	$\text{median}_r \hat{\theta}_r - \theta_0 $
BIAS(SE):	bias of estimated standard error in %	$\frac{100}{R} \sum_{r=1}^R [\hat{\sigma}_r(\hat{\theta}_r) - \sigma(\hat{\theta})] / \sigma(\hat{\theta})$

Note: $\hat{\theta}_r$ denotes the r th realization of the simulated estimator for θ . $\hat{\sigma}_r(\hat{\theta}_r)$ indicates the r th estimate of the standard errors of $\hat{\theta}$ based on the asymptotic standard errors. $\sigma(\hat{\theta})$ indicates the standard errors computed from the Monte Carlo replications.

Table 2: Simulation Results for Pure Random Effects ($\alpha = 0, \delta = \sqrt{0.5}$)

	RMSE $\times 10$		MAE $\times 10$		bias (SE) in %	
	β_D	β_N	β_D	β_N	β_D	β_N
<i>N=100</i>	<i>T = 5, 1000 replications</i>					
ML-RE	1.09	1.22	0.69	0.78	-5.4	0.4
Infeasible GMM	1.09	1.29	0.73	0.84	0.2	-3.9
Pooled	1.21	1.34	0.75	0.85	-1.4	-0.3
GMM-IV(param)	1.35	1.62	0.83	0.93	-10.4	-14.3
GMM-WNP	1.14	1.29	0.76	0.86	-0.7	-0.3
GT-Score	1.24	1.24	0.84	0.97	82*	27*
GT-Score ⁺	1.22	1.30	0.73	0.97	-85*	-84*
GT-MD	1.37	1.37	1.01	0.91	-12.2	-9.1
GT-MD ⁺	1.21	1.31	0.75	1.00	-3.6	-4.3
<i>N=400</i>	<i>T = 5, 1000 replications</i>					
ML-RE	0.52	0.61	0.35	0.42	0.8	-2.2
Infeasible GMM	0.55	0.60	0.37	0.39	-0.5	2.4
Pooled	0.58	0.67	0.39	0.45	2.0	-1.0
GMM-IV(param)	0.55	0.62	0.38	0.42	-1.9	1.4
GMM-WNP	0.54	0.63	0.36	0.43	1.9	-0.6
GT-Score	0.58	0.67	0.36	0.47	215*	108*
GT-Score ⁺	0.62	0.67	0.41	0.46	-87*	-86*
GT-MD	0.66	0.75	0.41	0.50	9.2	8.5
GT-MD ⁺	0.58	0.65	0.37	0.43	-5.7	-6.1
<i>N=100</i>	<i>T = 10, 500 replications</i>					
ML-RE	0.74	0.85	0.52	0.53	1.2	5.7
Infeasible GMM	0.80	0.89	0.55	0.61	-4.8	1.1
Pooled	0.86	0.94	0.58	0.61	-0.5	5.0
GMM-IV(param)	2.72	3.71	1.38	1.54	4458	739
GMM-WNP	0.84	0.96	0.58	0.60	-2.4	1.2
GT-Score	0.82	0.99	0.56	0.70	16.3*	-11.2*
GT-Score ⁺	0.80	1.00	0.53	0.64	-85.0	-86.1
GT-MD	0.93	1.13	0.66	0.71	8.9	-5.4
GT-MD ⁺	0.78	0.98	0.53	0.62	-0.7	-8.2
<i>N=400</i>	<i>T = 10, 500 replications</i>					
ML-RE	0.37	0.45	0.24	0.30	1.4	-0.3
Infeasible GMM	0.37	0.45	0.24	0.30	1.4	-0.3
Pooled	0.41	0.49	0.28	0.34	3.9	0.2
GMM-IV(param)	0.58	0.77	0.34	0.40	-19	-27
GMM-WNP	0.40	0.47	0.26	0.32	-1.4	-1.7
GT-Score	0.41	0.46	0.27	0.29	159*	129*
GT-Score ⁺	0.43	0.47	0.27	0.31	-86*	-85*
GT-MD	0.46	0.48	0.30	0.32	-12	-8.9
GT-MD ⁺	0.45	0.42	0.26	0.28	-5.8	4.6

Note: * The estimates of the standard errors are based on the assumption of a correctly specified model.

Table 3: Simulation Results for AR(1) and random effects ($\alpha = 0.8, \delta = 0.2$)

	RMSE $\times 10$		MAE	$\times 10$	bias (SE) in %	
$N=400$	$T = 5, 1000$ replications					
ML-RE	0.61	0.64	0.40	0.44	4.0	2.2
Infeasible GMM	0.61	0.56	0.41	0.45	−0.2	−1.3
Pooled	0.70	0.73	0.47	0.51	−1.3	0.9
GMM-IV(param)	0.63	0.68	0.42	0.47	−0.2	−1.7
GMM-WNP	0.61	0.68	0.41	0.46	2.2	−0.3
GT-Score	0.68	0.71	0.47	0.47	472*	388*
GT-Score ⁺	0.68	0.70	0.45	0.47	−85*	−85*
GT-MD	0.68	0.71	0.45	0.47	−22	−19
GT-MD ⁺	0.60	0.64	0.41	0.40	−8.3	−4.6

Note: * The estimates of the standard errors are based on the assumption of a correctly specified model.

Table 4: Descriptive Statistics

		mean	std.dev.
ln(sales)	ln of industry sales in DM	10.54	1.0
relative firm size	ratio of employment in business unit to employment in industry (* 30)	0.074	0.29
import share	ratio of industry imports to (industry sales + imports)	0.25	0.12
FDI-share	ratio of industry foreign direct investment to (industry sales + imports)	0.046	0.047
productivity	ratio of industry value added to industry employment	0.090	0.033
raw material	= 1 if firm is in sector 'raw materials'	0.087	0.28
investment	= 1 if firm is in sector 'investment goods'	0.50	0.50
dependent variable	= 1 if product innovation is realized	0.60	0.49

Table 5: Estimation Results for the Innovation Probit

	<i>pooled</i>	<i>GMM-WNP</i>	<i>GMM-IV(param)</i>	<i>GT-MD</i>	<i>GT-MD+</i>
ln(sales)	0.18 (0.04)	0.15 (0.04)	0.17 (0.04)	0.15 (0.06)	0.17 (0.03)
relative firm size	1.07 (0.31)	0.95 (0.20)	1.20 (0.31)	0.90 (0.21)	0.95 (0.20)
import share	1.13 (0.24)	1.14 (0.24)	1.12 (0.24)	0.99 (0.29)	1.05 (0.21)
FDI share	2.85 (0.68)	2.59 (0.59)	2.78 (0.68)	2.87 (0.79)	2.88 (0.53)
productivity	-2.34 (1.32)	-1.91 (0.82)	-2.54 (1.24)	-2.49 (1.67)	-2.50 (0.70)
raw material	-0.28 (0.13)	-0.28 (0.12)	-0.24 (0.13)	-0.20 (0.11)	-0.24 (0.07)
investment	0.19 (0.06)	0.21 (0.06)	0.19 (0.06)	0.20 (0.08)	0.20 (0.05)
constant	-1.96 (0.38)	-1.74 (0.37)	-1.86 (0.37)	-1.68 (0.54)	-1.90 (0.31)

Table 6 Estimated Correlation Matrix

	1985	1986	1987	1988
1984	0.60	0.65	0.58	0.48
1985		0.64	0.56	0.37
1986			0.68	0.58
1987				0.64